Neurofeedback Parameters Meta-Analysis

2024-05-22

Summary

A meta-analysis was conducted testing the influence of different neurofeedback (NF) training parameters on changes in neural activity from (1) first NF training session to last training session (k = 35) and (2) pre-training baseline to post-training rest (k = 20). Analyses were conducted on the following predictors: training duration, presence of pretraining rehearsal, imaging device, quality assessment score, presence of additional training instruction, blinding, functional localizer, year and age. Predictors with too few cases per level (<10) were excluded from the analysis.

Methods

- **Convert SE to SD**: All pre and post variability measures were converted to standard deviation in order to calculate the standardized mean difference (SMD)
- The **standardized mean difference (SMD)** was calculated using Hedge's g, as this is appropriate for studies with small sample sizes (Mean N = 16.23, SD = 12.67) (Hedge, 1981). This was done separately for last vs. first training session and post-training vs. baseline.
- **Adjust for Direction of Effect**: In studies where the task was to decrease activity, the SMD was multiplied by -1 to ensure that a positive SMD indicates a change in the predicted direction.
- **Pooled effect size**: Random effects model was used to calculate the pooled effect size. The Restricted Maximum Likelihood (REML) method was used for estimating the between-study variance (tau^2). REML was used since it produces results with reduced bias (Veroniki et al., 2016). The R package meta was used to calculate the pooled effect sizes (Balduzzi, 2019).
- **Outlier detection**: Graphic Display of Heterogeneity (GOSH) plot analysis (Olkin, 2012) was performed in order to identify influential cases which had a significant contrbution to heterogeneity.

Meta-regression and Sub-Group Analyses:

- **Check for Multicollinearity**: Correlations between predictor variables were assessed to check for multicollinearity. Variables with correlations ≥ 0.8 were considered for removal. No significant multicollinearity was detected.
- **Multi-model inference**: was used to identify the best-fitting model and the most influential predictors (Harrer et al., 2009).
- **Meta-Regression Analysis**: Meta-regression was performed to investigate the effect of potential moderators on the effect sizes.

- The following **sub-group analyses** were performed: blinding, imaging device, instruction provided, presence of pre-training rehearsal, functional localizer.

Results

- When testing the overall difference between the first and last training sessions there was a significant effect of training (g = 0.34, t = 6.31, p < 0.001).
- There was no significant effect when testing the difference between pre-training baseline and post-training activity (g = 0.2, t = 2.05, p = 0.056).
- The effect size for studies that used EEG was significantly higher (g = .40) than those that used fMRI (g = .19) (p = .01)
- The effect size for studies that included a pre-training rehearsal trial was significantly lower (g = 0.20) than those that did not include one (g = 0.40) (p = .04)
- Although multimodel inference analysis classified training duration as the most important predictor, there was no significant relationship between standardized mean differences and training duration(I^2 = 71.51%, R^2 = .77%, $F(1,32)$ = 3.32, p = 0.078)

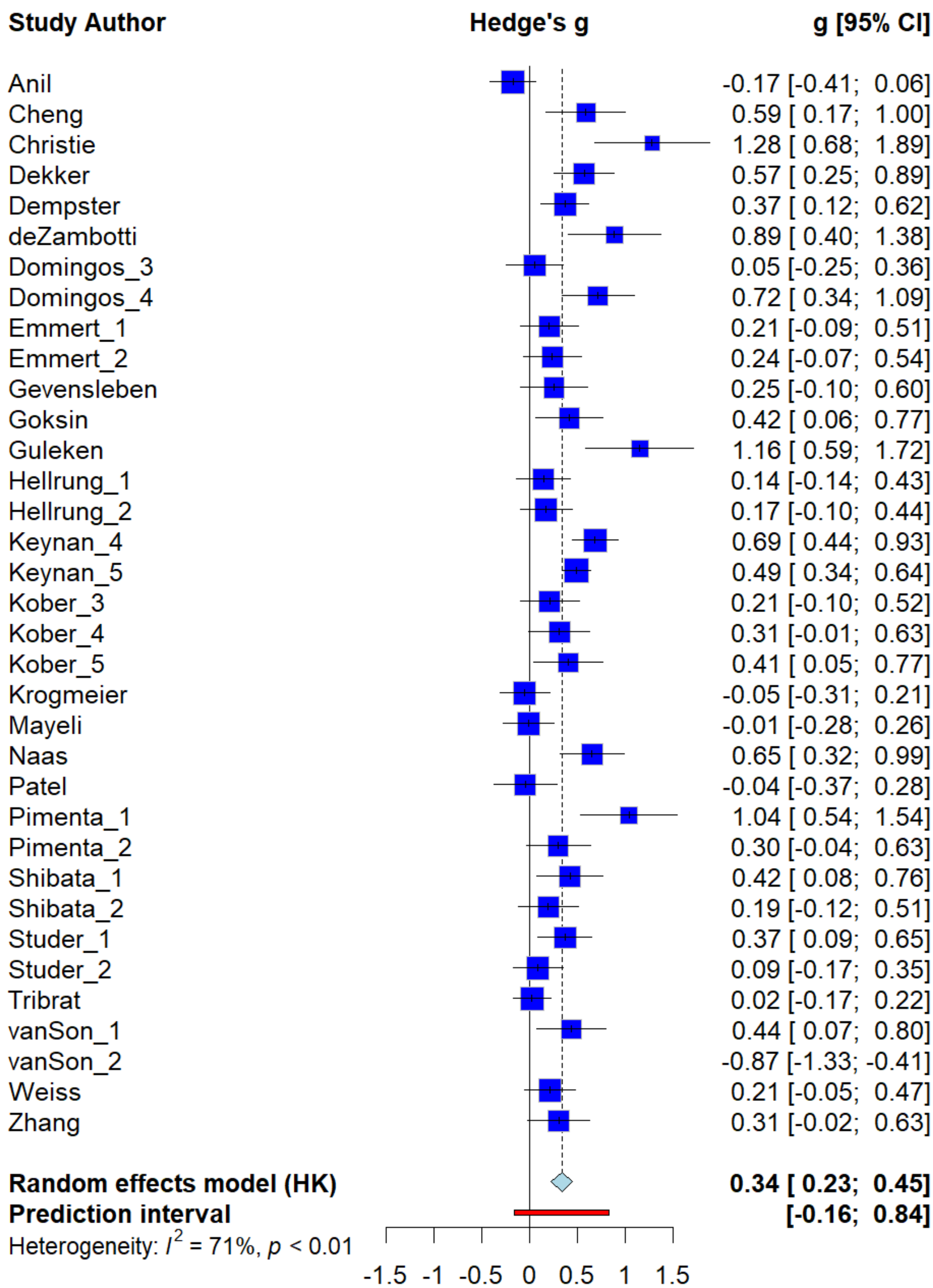
Next Steps and Questions

- Go back to the papers that reported more than one type of feedback and maybe try to find a way to further aggregate studies or reduce categories.
- Regions of interest might need to be aggregated into categories. Discuss with group on how to do this.
- Need to make a comparison between different targeted frequency bands. Some frequency bands have too few studies and might need to be excluded. Another option is to treat frequency as a continous variable.
- Plot risk of bias
- How can we include data from multiple sessions? One option could be to run a regression between SMD and time since first session with amount of training up to that point as a covariate. Open to suggestions.
- How do we apply correction for multiple comparisons here?
- Need to clean methods and results for the manuscript. Also not sure what are all the values that need to be included with each result.

Overall Effect of Training

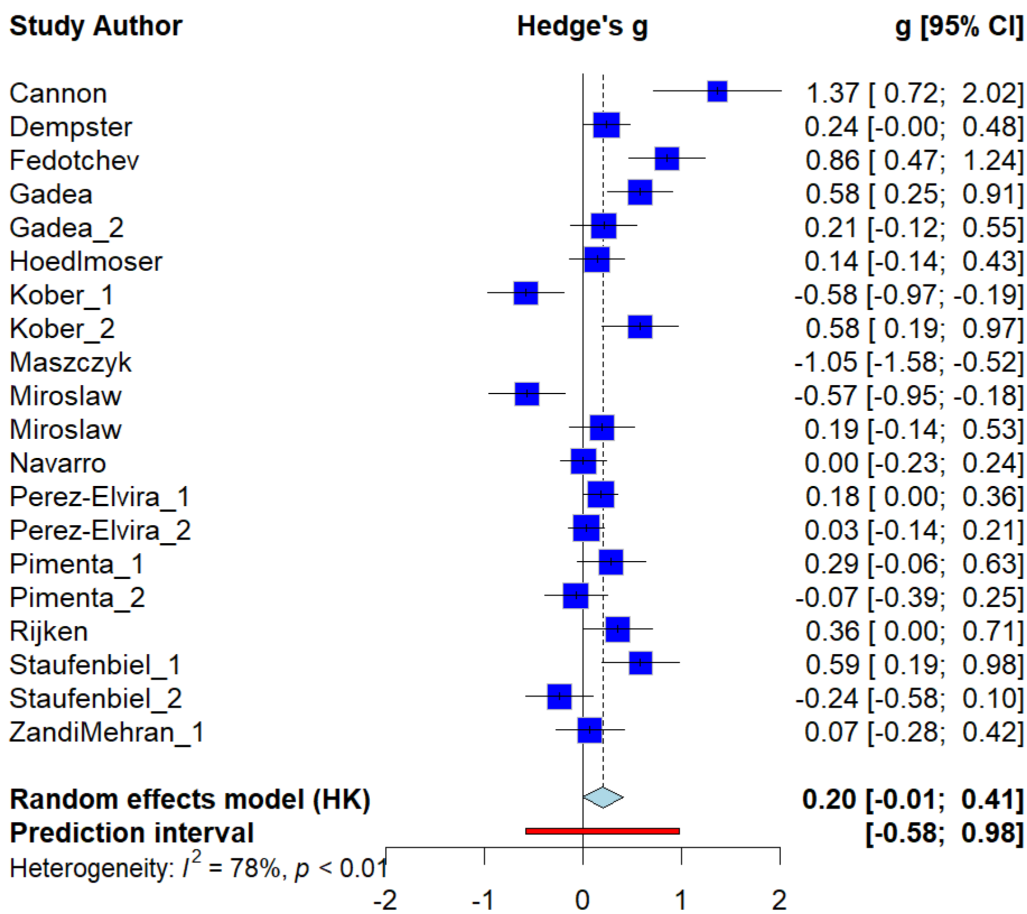
A random effects model was used to test the overall effect of training on neural activity.

Last vs. First training session



When testing the overall difference between the first and last training sessions there was a significant effect of training (Hedge's g = 0.3369, CI = [0.2284; 0.4455], t = 6.31, p < 0.001)

Post-training vs. Pre-training Baseline



There was no significant effect of training when testing the difference between pre-training baseline and post-training activity (g = 0.2, t = 2.05, p = 0.056).

Sensitivity Analyses on Outlier Influence

Outliers were detected using the *graphic display of heterogeneity (GOSH)* plot analysis. Sensitivity analyses comparing the results before and after removing the outliers are reported below.

Last vs. First training session

Analysis	*g*	*95%CI*	*p*	*95%PI*	*I2*	*95%CI*
Main Analysis	0.32	0.19-0.44	<0.001	-0.29-.93	75.4%	65.9-82.2
Infl. Cases Removed[1]	0.37	0.23-0.45	<0.001	-0.16-0.84	70.9%	58.8-79.4

[1] Removed as outliers: Van Son (group B).

Post-training vs. Pre-training Baseline

Analysis	***g***	***95%CI***	***p***	***95%PI***	***I2***	***95%CI***
Main Analysis	0.15	-0.08-0.38	0.19	-0.78-1.08	81.2%	71.8%-87.4%]
Infl. Cases Removed[2]	0.2	0.005-0.40	0.056	-0.58-0.98	77.9%	66.0%-85.7%

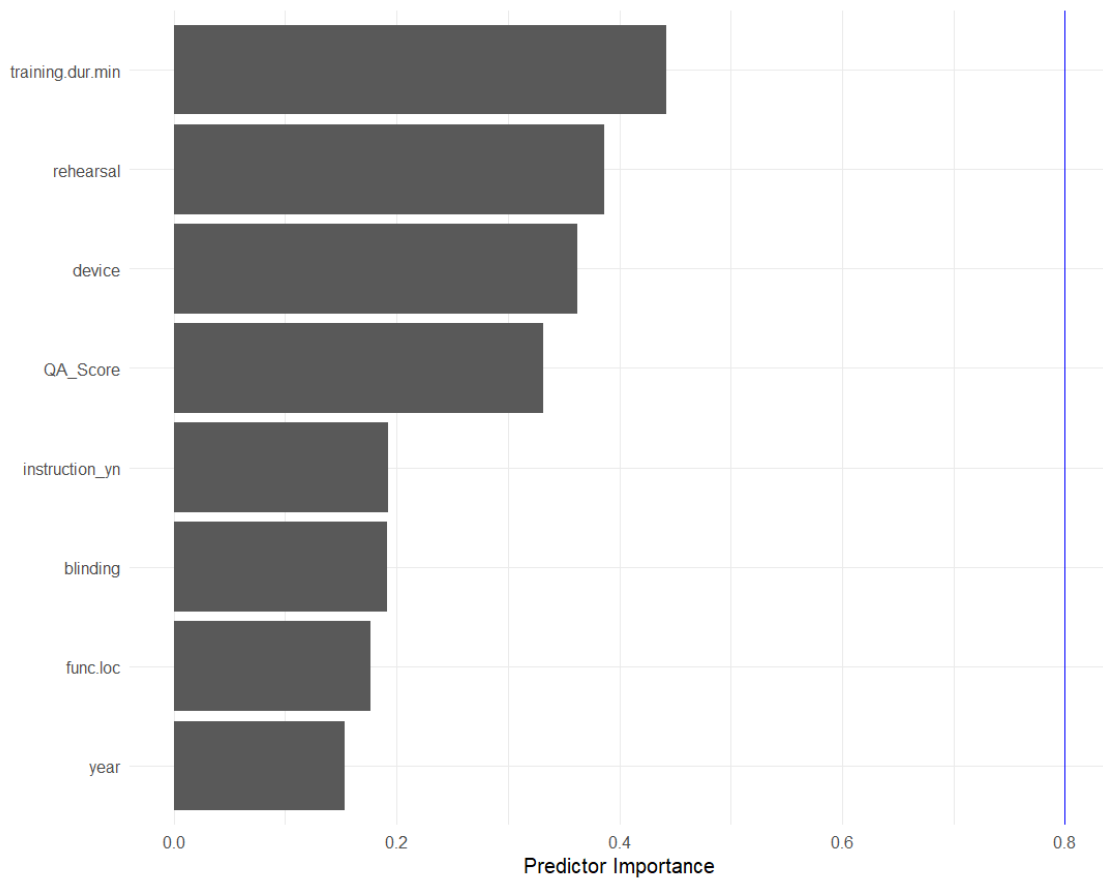
[2] Removed as outliers: Maszczyk.

Meta-Regression

Multimodel Inference

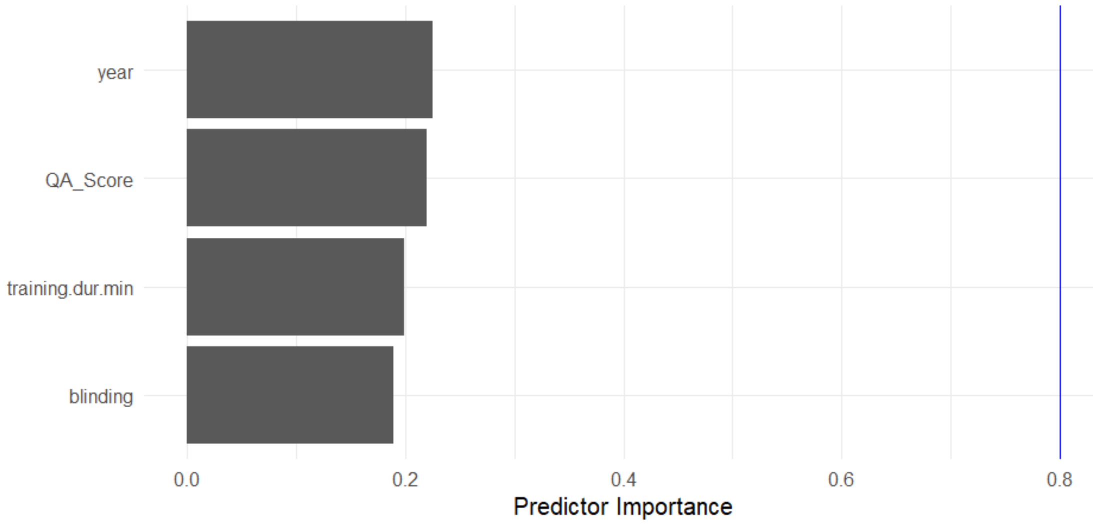
Multimodel inference was used to identify the best-fitting model and the most influential predictors

Last vs. First training session



Predictors in Last vs. First training session ranked by importance. Results from the analysis indicated that best fit occurs when predictors are looked at separetely

Post-training vs. Pre-training Baseline

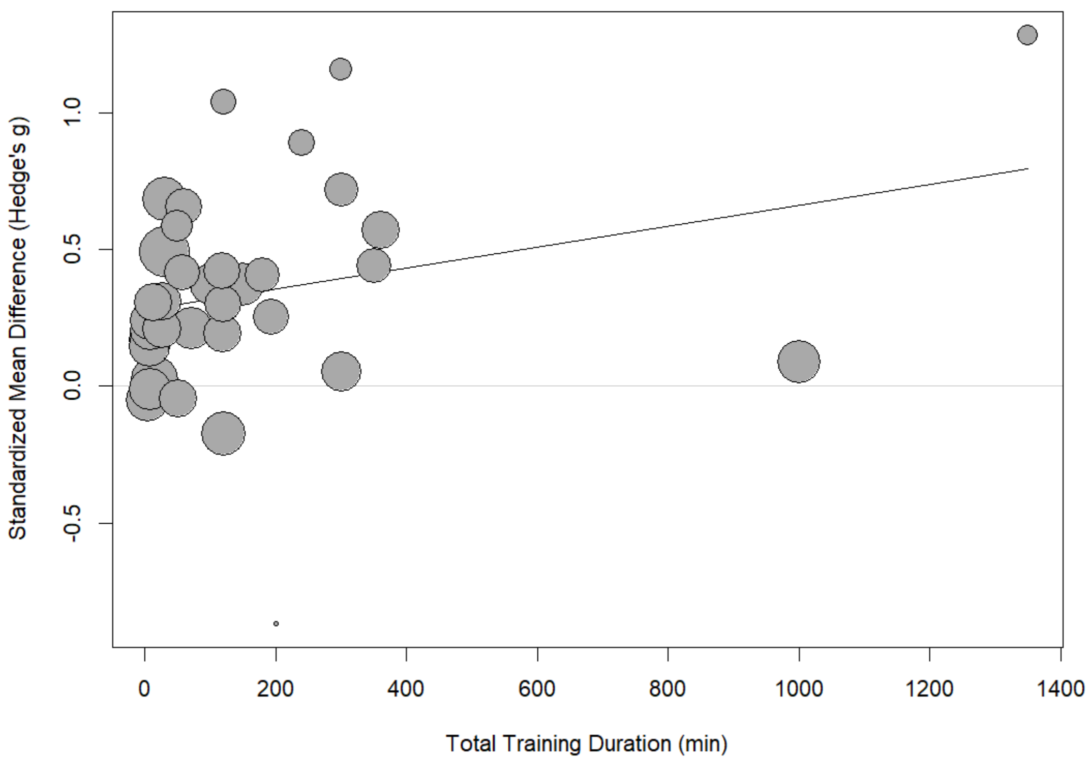


Predictors in Post-training vs. Pre-training Baseline ranked by importance. Results from the analysis indicate that best fit occurs when predictors are looked at separetely

Mixed-Effects Model Regression

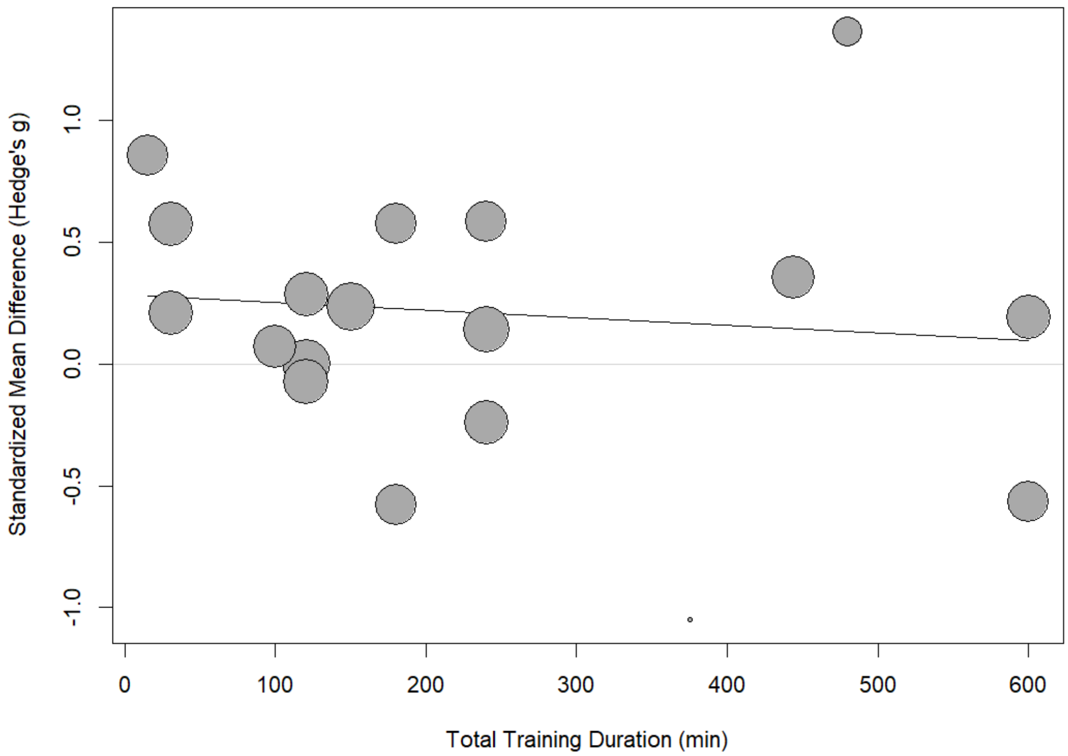
Training Duration

Last vs. First training session



There was no significant relationship between Standardized Mean Difference (Hedge's g) and total training duration. (k = 34, I^2 = 71.51%, R^2 = .77%, $F_{(1,32)}$ = 3.32, p = 0.078)

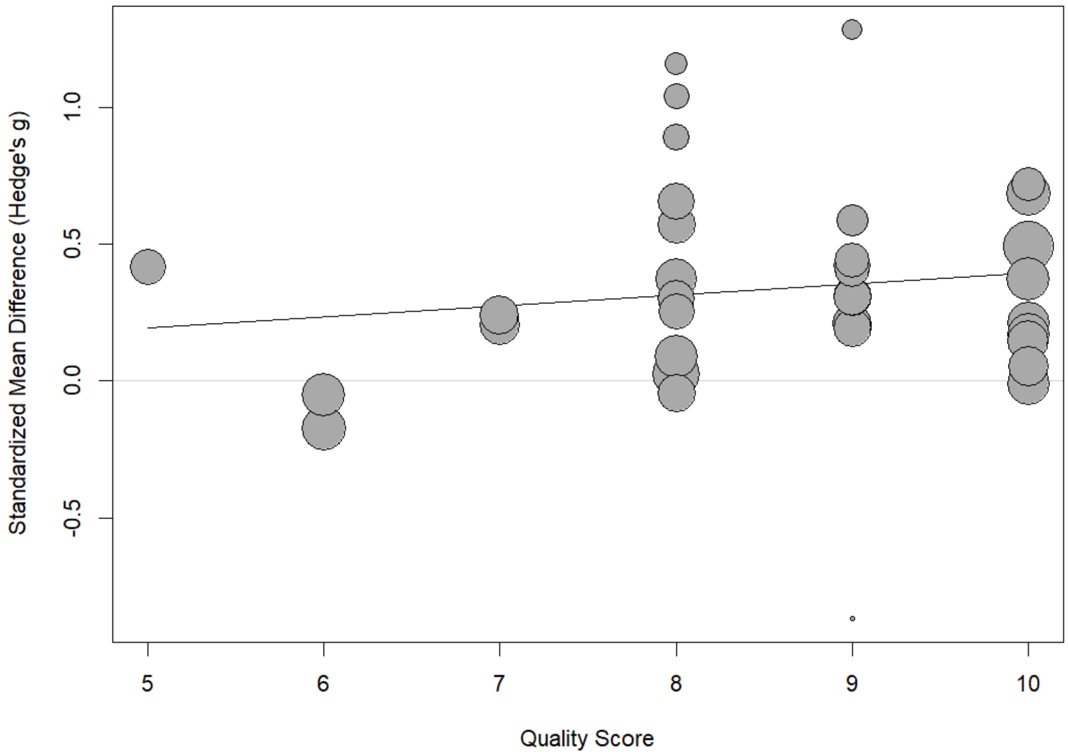
Post-training vs. Pre-training Baseline



There was no significant relationship between Standardized Mean Difference (Hedge's g) and total training duration. (k = 17, I^2 = 84.98%, R^2 = 0.0%, $F_{(1,15)}$ = 0.24, p = 0.63)

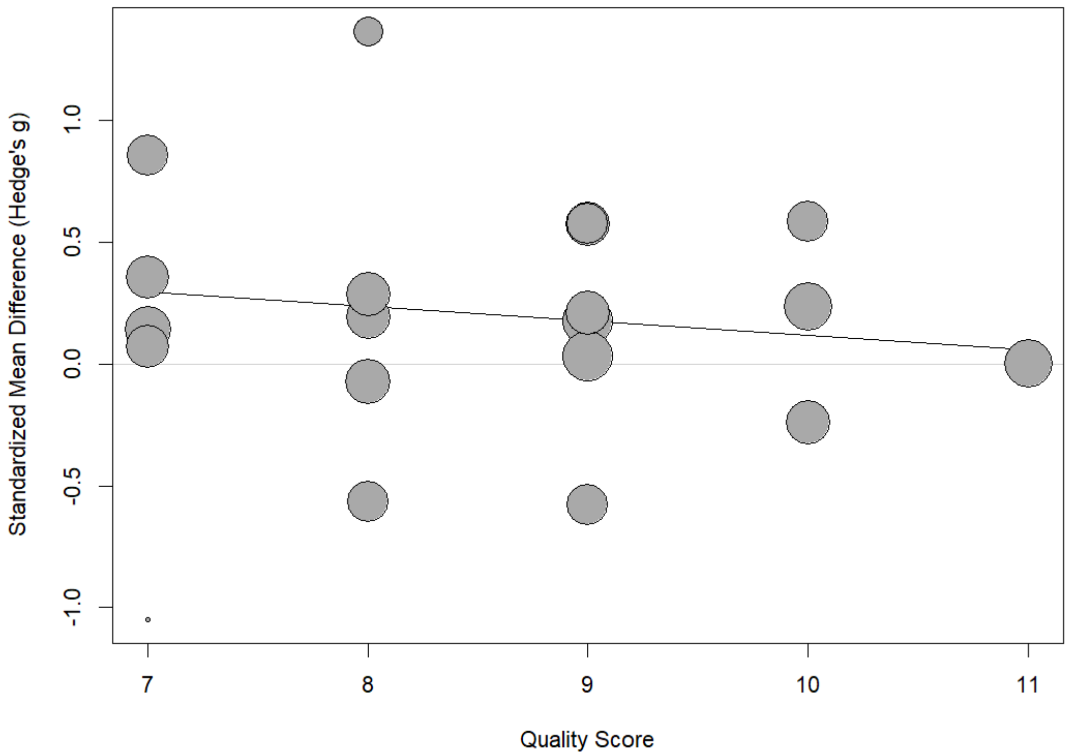
Quality Score

Last vs. First training session



There was no significant relationship between Standardized Mean Difference (Hedge's g) and Study Quality Score. (k = 34, I^2 = 70.77%, R^2 = 3.0%, $F_{(1,32)}$ = .96, p = 0.34)

Post-training vs. Pre-training Baseline



There was no significant relationship between Standardized Mean Difference (Hedge's g) and Study Quality Score. (k = 19, I^2 = 85.24%, R^2 = 0.0%, $F_{(1,17)}$ = 0.48, p = 0.50)

Age

Last vs. First training session

There was no relationship between age and SMD (k = 23, I^2 = 64.9%, R^2 = 5.13%, $F_{(1,21)}$ = 1.45, p = 0.24)

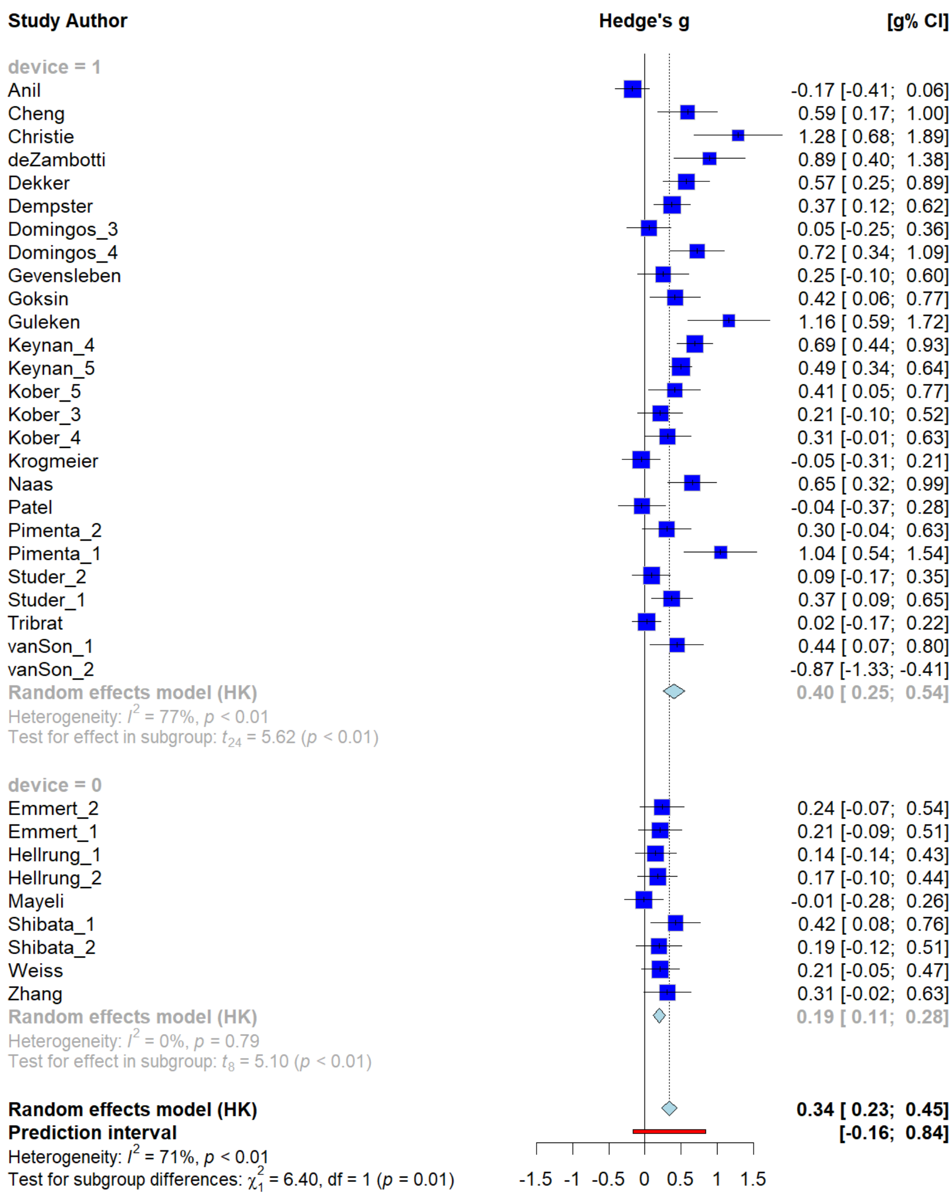
Post-training vs. Pre-training Baseline

There was no relationship between age and SMD (k = 13, I^2 = 87.01%, R^2 = 0.0%, $F_{(1,11)}$ = 0.42, p = 0.53)

Subgroup Analyses

Imaging Device Used

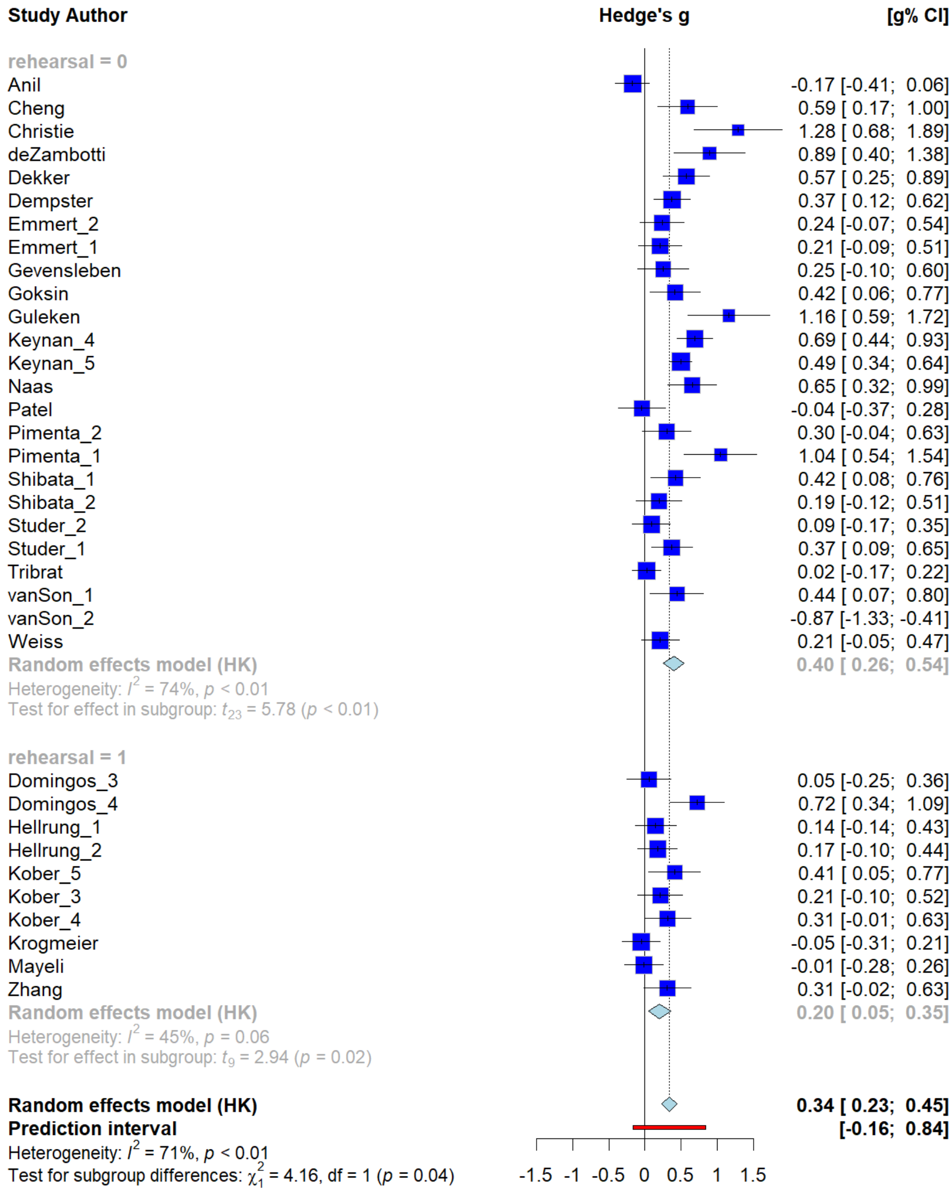
Last vs. First training session



Subgroup analysis comparing EEG (device = 1) and fMRI (device = 0) studies' effect sizes. The effect size for studies that used EEG (g = .40) was significantly higher than those used fMRI (g = .19) (p = .01)

Pre-training Rehearsal Present

Last vs. First training session



Subgroup analysis comparing studies with pre-training rehearsal (1) and no rehearsal (0). The effect size for studies that included a pre-training rehearsal trial was significantly lower (g = 0.20) than those that did not include one (g = 0.40) (p = .04)

Blinding

Last vs. First training session

There was no difference between blinding (k = 15, g = 0.35) and no blinding (k = 19, g = 0.33) (Q = 0.05, p = .82)

Post-training vs. Pre-training Baseline

There was no difference between blinding (k = 9, g = 0.25) and no blinding (k = 10, g = 0.16) (Q = 0.15, p = .70)

Functional Localizer Used

Last vs. First training session

There was no difference between the presence of a functional localizer (k = 10, g = 0.29) vs non (k = 24, g = .36) (Q = 0.53, p = .47).

Presence of Instruction for Feedback Training

Last vs. First training session

There was no difference between the presence of a additional instruction (k = 12, g = 0.26) vs non (k = 22, g = .37) (Q = 1.15, p = .28).